Utilizing Sentinel-1A Radar Images for Large-Area Land Cover Mapping with Machine-learning Methods

Imangholiloo, Mohammad

2019-03

Imangholiloo , M , Rasinmaki , J , Rauste , Y & Holopainen , M 2019 , ' Utilizing Sentinel-1A Radar Images for Large-Area Land Cover Mapping with Machine-learning Methods ' , Canadian journal of remote sensing , vol. 45 , no. 2 , pp. 163-175 . https://doi.org/10.1080/07038992.2019.1635877

http://hdl.handle.net/10138/324454
https://doi.org/10.1080/07038992.2019.1635877

unspecified
acceptedVersion

Utilizing Sentinel-1A Radar Images for Large-Area Land Cover Mapping with Machine Learning Methods

Mohammad Imangholiloo[a]*, Jussi Rasinmäki[b], Yrjö Rauste[c], Markus Holopainen[a]

[a] Department of Forest Sciences, University of Helsinki, Helsinki, Finland; [b]Simosol Oy (Ltd.), Riihimäki, Finland; [c] Big data industrial applications, VTT Technical Research Center of Finland, Espoo, Finland

* Corresponding author. E-mail address: Mohammad.Imangholiloo@helsinki.fi.

Utilizing Sentinel-1A Radar Images for Large-Area Land Cover Mapping with Machine Learning Methods

Land use and land cover maps are vital sources of information for many applications. Recently, using high-resolution and open-access satellite images have become a preferred method for mapping land cover especially over large areas. This study was designed to map the land cover and agricultural fields of a large-area using Sentinel-1A synthetic aperture radar (SAR) images. Seven machine learning methods were employed for image analyses. The Random Forest classifier algorithm outperformed the other machine learning methods in the training step, thus we selected for further use and tuned its parameters. After several image processing steps, we classified the final image into 23 land cover classes and achieved an overall accuracy of 42% for all classes, and 57% for agricultural fields. This research note highlights some characteristics of Sentinel-1A images and provides novel methods for nation-wide large-area mapping applications. The results demonstrate the potential advantages of using Sentinel-1 images for land cover mapping.

Keywords: Sentinel-1A; SAR; land use mapping, land cover mapping; machine learning; Random Forest

Utilisation des images radar Sentinel-1A pour la cartographie de la couverture terrestre sur des grandes étendues à l'aide de méthodes d'apprentissage automatique

Les cartes qui présentent les aspets d'utilisation et d'occupation des terres sont des sources d'informations cruciales pour des nombreuses applications. Aujourd'hui, l'utilisation des images satellitaires à haute résolution spatiale et à accès ouvert est devenue une méthode privilégiée pour cartographier la couverture terrestre, en particulier sur de vastes étendues. Cette étude avait comme objectif principal; cartographier la couverture terrestre et les champs agricoles d'une vaste étendue à l'aide des images radar à synthèse d'ouverture (SAR) Sentinel-1A. Sept méthodes d'apprentissage automatique ont été utilisées pour l'analyse des images. L'algorithme du classificateur Forêt Aléatoire a surperformé les autres méthodes d'apprentissage automatique à l'étape de la formation; ainsi, nous l'avons sélectionné pour une utilisation ultérieure et réglé ses paramètres. Après plusieurs

étapes de traitement des images, nous avons classé l'image finale en 23 classes de couverture terrestre et avons atteint une précision globale de 42% pour toutes les classes, et de 57% pour les champs agricoles. Cette note de recherche mets en évidence certaines caractéristiques des images Sentinel-1A et fournit des nouvelles méthodes pour les applications cartographiques des étendues vastes à l'échelle nationale. Les résultats démontrent les avantages potentiels de l'utilisation des images Sentinel-1 pour la cartographie de la couverture terrestre.

Mots clés: Sentinel-1A; SAR; cartographie de l'utilisation des terres, cartographie de la couverture terrestre; apprentissage automatique; Forêt Aléatoire.

1. Introduction

Land use and land cover maps are important in the management of natural resources, forests, and agricultural fields, as well as other subjects. The current development and availability of open-access remotely-sensed data have made mapping processes less costly.

Synthetic Aperture Radar (SAR) has a history of over 25 years starting with ERS-1 satellite in early 1990s. Sentinel-1 is a continuation of the long history of SAR data utilization which comprises a constellation of two satellites: Sentinel-1A and Sentinel-1B. These satellites were launched by the European Space Agency (ESA) in April 2014 and April 2016, respectively. Both satellites carry C-band radar and are technically identical.

SAR images from Sentinel-1 provide data on land and sea areas and nearly full coverage of Earth, except for the polar regions. Sentinel-1 images are open data and have high temporal and spatial resolutions. Such advantages provide several possibilities to scientists and experts in remote sensing and Earth observation sciences, as well as many other fields. The SAR system is independent of cloud conditions and the sun light because radar can penetrate clouds and because radar uses its own source of radiation. High temporal data coverage provided by Sentinel-1A and Sentinel-1B will

eventually enable us compute multitemporal metrics, especially those for cloud-covered areas (Balzter et al. 2015).

The open-access and high-resolution Sentinel-1 products have been utilized in several land cover and agricultural cropping studies—for example, CORINE land cover mapping (Balzter et al. 2015), mapping rice and cropping schemes (Nguyen et al. 2016), cropping-system diversity (Dimov et al. 2016), and urban mapping (Tapete and Cigna 2016; Abdikan et al. 2016; Alexandrer and Ban 2015; Haas and Ban 2015). More specifically, Nguyen et al. (2016) delineated rice-cultivated areas using Sentinel-1 and a decision-tree approach that included seasonal phenological parameters and a time-series analysis. Moreover, Dimov et al. (2016) mapped the diversity in three cropping systems using Random Forest algorithm using Landsat-8 and Sentinel-1. In addition, Balzter et al. (2015) used Sentinel-1A data to map CORINE land cover into 17 land cover classes.

The Random Forest classifier has been reported as a promising method for land cover mapping (Pal (2003); Gislason et al. (2006); Rodriguez-Galiano et al. (2012a); Rodriguez-Galiano et al. (2012b)). For example, Pal (2003) obtained a classification accuracy of 88.32% by using the Random Forest classifier on Landsat-7 ETM^{+} images for 7 land-cover classes. Moreover, he claimed that the Random Forest algorithm maintains a reliable performance if noise exists in the training data. In line with Pal (2003), Gislason et al. (2006) revealed that the Random Forest classifier is faster in training, requires less guidance, and can estimate the importance of variables for the classification and detection of outliers. Rodriguez-Galiano et al. (2012a) also confirms that the Random Forest algorithm can reduce training data and noise.

To the best of authors' knowledge, the Sentinel-1 data has not been applied for land-cover mapping of nation-wide large areas, especially for detailed classification (23 thematical land cover classes). Thus, this study was designed to analyze the applicability of Sentinel-1A (SAR) images for nation-wide large-area mapping, and to

discuss about characteristics of the Sentinel-1A images with proposing novel methods such as mosaicking many SAR images. This study also investigates to what extent agricultural crop types can be separated from one another and from other land cover classes.

2. Material and Methods

2.1. Study area

The study area covered 476,672.4 km^2, mainly over Indus Valley in Pakistan (Fig. 1). The farming activity is regulated in relation to monsoon rains, which occur from June through September (Imran et al. 2014). There are two cropping seasons in the study area. The rain pattern varies in time and space; that is, where and how much it rains can vary in every season in a given year (Imran et al. 2014). The land area is used for farming crops such as rice, wheat, cotton, maize, and corn, and it includes orchards and tree plantations with a few natural classes of vegetation.

2.2. Materials

2.2.1. Remote sensing data

We downloaded Ground Range Detected High-Resolution (GRDH, VV-polarization) images acquired in interferometric wide swath (IW) mode, from the European Space Agency's (ESA) Sentinels Scientific Data Hub server. The GRDH product had a ground range geometry, processed in level 1, and had a higher resolution than other major operator modes over land. Therefore, it was regarded as the best choice for land use and land cover mapping and monitoring. It was determined that it would be better if all data were acquired at the same incidence angle. In Sentinel-1/IW images, the incidence angle varies over 10 degrees, from the near range to the far range. In a country-wide monitoring application, it would be difficult or impossible to stratify each image swath

into 3 (or more) strata based on the (nominal) incidence angle (or beam swath). We could only accept that the incidence angle variation is one of the factors that would contribute to the errors of the monitoring algorithm.

As shown in Table 1, at least 33 images from Sentinel-1 were required to fully cover the study area. The images (Fig. 2) were mosaicked and referred to as a feature (epoch or raster bands)—that is, a time-limited set of images that covered the analysis area and were mosaicked after being processed.

An experiment was conducted to find optimal times for the features in the Sentinel-1 time-series. As the results were not clear, it was decided that the times of features are adapted to the times of features used in an earlier Landsat-8 study (World Bank 2016). Because Sentinel-1 data was not available before October 2014, corresponding time from year 2015 was used instead (Table 1 and Table 2), even though the image acquisition time did not completely overlap with the time period of ground data collection. New ground data collection for 2015 was not possible in this study. The crop types may have changed from 2014 to 2015 on some farms, and this forms an additional source of error in this study.

The image downloading process was automated by a Simosol-made script (early version available on Github[1]). The script acquired a date (we called it the "target date") to find a full coverage of Sentinel-1A images over a given analysis area. The middle of the date range of Landsat-8 (Table 2) was calculated and regarded as the "target date" as well. The date range of images that were picked from the Sentinel-1A products (Table 1) was reported by the download script. The date ranges were compared and controlled to see whether the found images had any overlap—that is, whether an image was included in two features. If a time overlap was observed between the found images per epoch, we

[1] https://github.com/brenogil/scihub

slightly adjusted the target date to solve such issues.

The selected images' percentage of coverage of the study area was reported by the script after reporting the candidate images to download. The percentage was then double checked by the user to be more than 99.99%. The download step was conducted by giving the target date and by saving images in the separate subdirectories, ultimately to process and mosaic the images of each target date separately. Next, the downloaded and processed images of each time limit were mosaicked, and the feature (epoch) was created in this step. The features were then merged (layer-stacked) to be used in next steps.

2.2.2. *Field data*

The field observation data that were used in this study had been formerly collected and employed in the World Bank (2016) study. The agricultural classes were collected with MHG's (MHG Systems Oy Ltd., Finland) mobile software during the field survey by asking farmers about which crops had been cultivated in the current and previous cropping seasons (World Bank 2016). Next, the data were processed into "per cropping season reference observations" (World Bank 2016). The dash line between crop names (Table 3 and Table 4) in the field classes column refers to the cultivated crop type in the first and second cropping season (within a year). The data of each observation had geographic coordinates and information on the cultivated crop. A separate quality-control procedure was executed at this stage to select only those samples that clearly represented a field pixel (World Bank 2016). This field observation provided reference data with a sample size of 3,161 observations.

The field data for other land-cover classes, referred to as "other classes" (Table 3), were collected with very high-resolution imagery, which was available online and provided 2,705 reference samples. These land use classes were generated by randomly created image samples, with an approximate area of 300 m × 300 m, at a 50-cm spatial

resolution. Next, the images were visually interpreted and assigned to the following land use classes: natural vegetation (high, medium, or low biomass), urban area, road, and water. The high, medium, and low biomass classes (Table 3) denoted natural vegetation or woody biomass cultivation areas, such as orchards and tree plantations.

Table 4 shows the number of field observation data used for training (ground reference) and validation per land cover type.

2.3. Methodology

2.3.1. *Processing the SAR data*

The processing chain of SAR data was created with the graph processing tool (GPT) in SNAP[2] desktop (version 3). Next, we saved the graph as an xml file and ran it in the Linux server over batches of images automatically.

The multilook operator were used to create square-shaped pixels, reduce noise, and improve radiometric resolution but at the cost of decreasing spatial resolution, because of aggregating pixels (Kilsedar 2016; Veci 2015). In the present study, the research team decided to perform a multilook process in 2 × 2 (number of range looks × number of azimuth looks) with "independent looks" to achieve output images with a 20 m × 20 m pixel size. The pixel size of 20 m was used to reduce noise in the SAR data, to fasten the processing time of the next operators, and to lower data size.

The SAR images in the calibration operator were radiometrically corrected and calibrated; meaning that, pixel values became the true radar backscatter value of the reflecting surface. The calibrations that were applied in this operator were "mission specific," and the SNAP software automatically defined the type of calibrations on the

[2] SNAP - ESA Sentinel Application Platform version 3, http://step.esa.int

196 basis of the metadata of the input (Kilsedar 2016; Veci 2015).

197 In the terrain-correction step, the images were geocoded by correcting SAR geometric
198 distortions (Kilsedar 2016; Veci 2015). SRTM 1Sec HGT DEM maps were used for the
199 correction. The SNAP software downloaded the selected DEM from the Internet server
200 automatically when running the process.

201 The "terrain flattening" operator was used to reduce the effect of topography on the
202 radiometry of the SAR images. The SNAP program produced the output images in the
203 Plate-Carre (regular grid in latitude and longitude) projection. Images were then
204 reprojected to UTM zone 42 with GDAL[3] (gdalwarp utility). UTM zone 42
205 North/World Geodetic System 1984 was chosen because UTM 42N was used in the
206 middle of the Pakistan. Bilinear interpolation was used in all resampling steps in SNAP
207 and GDAL.

208 2.3.2. *Removing noisy marginal pixels*

209 The Sentinel-1A images had noisy marginal pixels with abnormally low pixel values
210 that should be treated before image classification. The noisy pixels can be grouped into
211 two classes: (*i*) pixels in the far and near ranges, which is the west and east sides of the
212 descending orbit images, and (*ii*) pixels in the north and south part of images. The noisy
213 pixels of class *i* were treated by developing an entirely novel method to automatically
214 clip parts of overlapping sides of images. Manually created per-orbit polygons (Fig. 3)
215 were used to identify class-*i* pixels and to set them as missing. The noisy pixels of class
216 *ii* were clipped by setting the pixels with value < 0.002 (clearly smaller than original
217 real pixels) as missing. Note that our novel method was required to remove the noisy
218 pixels of class *i*, because these noisy pixels had higher values.

[3] GDAL. 2016. Geospatial Data Abstraction Library (GDAL): Version 2.1.9, Open Source Geospatial Foundation, http://gdal.osgeo.org

219 2.3.3. *Image processing in GDAL*

220 A simplified flowchart of the GDAL-processing chain is shown in Fig. 4. Marginal
221 noisy pixels were clipped and set as missing (Section 2.3.2) before mosaicking images
222 to create features and changing the data types to integer 16 bit. Next, we used the
223 gdalwarp program to define output coordinate reference system, cell size, bounding box
224 (to get same pixel spacing for all layers) and nodata values that all were integrated in a
225 single command[4]. Finally, we merged the seven features using merge.

226 2.3.4. *Image Analysis*

227 Several machine-learning algorithms were tested in the modeling or training step—such
228 as AdaBoost, Gaussian Naive Bayes (GaussianNB), KNeighbors, linear discriminant
229 analysis, quadratic discriminant analysis, Support Vector Machine (SVC module), and
230 the Random Forest classifier.

231 The field data were split into two parts: training data (80%) and validation data (20%),
232 which was the same as used in World Bank (2016). The validation data (20%) remained
233 untouched and were used as an independent data set to evaluate the accuracy of the
234 prediction, but the training data (80%) was then split again into two components to be
235 used for training in the training or modeling step (80%) and to validate the training and
236 modeling (20%) of the models. In the modeling step, the scripts read the pixel values
237 located under the training data set and generated a model. The model was validated
238 using the proportion of training data relegated to validating the models (20% of the
239 80%) (Fig. 5).

240 We chose the Random Forest method for further use in this study and continued
241 experimenting to define optimal parameters for the Random Forest classifier. Simplified
242 and schematized machine-learning-based processing chain, used after GDAL steps, is

[4] gdalwarp -s_srs EPSG:4326 -t_srs EPSG:32642 -tr 20.0 20.0 -te -238580.0 2615430.0 1105480.0 3904909.0 -r bilinear -dstnodata -99 -cutline study_area.shp -of ERS input.tif output.ers

shown in Fig. 5.

The size of the 7-feature Sentinel-1 image was 60 GB with dimensions of 67,203 × 64,474 pixels. The memory limits of the used Linux server required that the image be split into three parts (northern, middle, and southern Pakistan) in the prediction phase. The computing of the prediction (classification) phase took 12 days. The training phase was not affected because a single model was applied to all three parts, which were later combined into a single land cover map. After testing different parameters of the Random Forest classifier, we used 5,000 trees for the number of estimators, entropy for criterion that controls and measures the quality of a split, and balanced weight that weights samples by the class frequency.

3. Results and discussion

3.1. Machine-learning image classification

As shown in Table 5, in the modelling phase, the Random Forest classifier outperformed the others with overall accuracy of 43%, then KNeighbors provided the second-best accuracy (37%) and lastly SVC method (14%). Accuracy of image classification phase yielded an overall accuracy of 42% for all classes and 57% for agricultural classes (Fig. 7). The maize-rice, rice-rice, rice-wheat and medium biomass classes achieved the higher accuracy within the 23 classes. The cotton-maize, cotton-sugarcane, maize-sugarcane, rice-sugarcane, and maize-wheat classes achieved the lower accuracy which were confusing with cotton-wheat, rice-wheat, and high biomass classes using the Random Forest classifier. Another confusion matrix was also created to analyze the agricultural classes within the same 23 classes classification. As expected, the accuracy was better for agricultural classes due to less classes (16 classes)

268 comparing to all classes (23 classes).

269 The results of the training step were correlated with the amount of training data,
270 especially in KNeighbors and the Random Forest classifier. Thus, land cover classes
271 with dominant training data were overestimated and rare classes underestimated, which
272 was in line with an earlier study that used Landsat-8 images for mapping the land use of
273 the same area (World Bank 2016). For example, cotton-rice and rice-sugarcane classes
274 had only 7 and 15 training points, respectively (Table 4). However, cotton-wheat and
275 rice-wheat classes had 944 and 735 training points, respectively. The numbers for the
276 validation data were 6, 7, 243, and 200, respectively. We note that the number of
277 observations for some land cover classes was too few for a meaningful Random Forest
278 classification (such as cotton-rice, cotton-other-crop, and maize-rice (Table 4)). Another
279 limitation was the imbalanced proportion of training and validation data in each class,
280 which could cause errors in classification without proper handling of imbalanced
281 training data. We shall note that the imbalance proportion of training data can reflect the
282 actual low or high presence of the class on the ground. We did not combine the small
283 classes into bigger ones, to enable comparing the results with the prior study (World
284 Bank 2016). In future studies, it is advisable to analyze handing imbalanced training
285 data, which was not our main focus of study.

286 The final land use map of Pakistan was created using Sentinel-1A images from 2014
287 and mainly 2015 with 239 scenes (Fig. 8). The novel Sentinel-1A satellite image and
288 method used in this study yielded as reliable results as in the earlier studies (World
289 Bank 2016; Balzter et al. 2015). Considering the number of classes, the performance of
290 this study was promising and comparable with Balzter et al. (2015) that used Sentinel-
291 1A images to map a CORINE land cover. However, the number of classes in our study
292 was higher (23) than in the study by Balzter et al (2015) with 17 classes. We achieved
293 an overall accuracy of 42% and they reported 68.4%. Moreover, Makinde and Oyelade

(2018) used dual-vertical Sentinel-1A and maximum likelihood classification method to map land cover of Lagos State for 4 land-cover classes. They achieved an overall accuracy of 76%. Considering the number of classes in the studies, the differences in the accuracy are justifiable. To achieve a better classification accuracy, it is recommendable to have maximum of 10–15 classes by discarding rare classes or fusing them together, for example combining cotton-rice and cotton-other crop from agricultural crops due to insufficient training data, or combining biomass classes together. Dual-polarized Sentinel-1 data is also recommended where available.

3.2. Limitations and solutions

There was a one-year time difference between field and satellite data among four features (Table 1), which could hinder achieving the higher accuracy. We assumed the presence of the same land use or land cover, especially the same cultivation in the year following the collection of the field data. This assumption was made to enable the use of field data of 2014 for 2015. In reality, this assumption did not necessarily represent all of the farms accurately because they might have changed the crops they cultivated.

Additionally, some 1–2-pixel wide gaps between the images along the orbit existed in the final image. This was unavoidable because the Sentinel-1 image boundaries were inconsistent along the orbit in some cycles (Fig. 9). Some interpolation methods could be employed to fill the gaps, but if it is not properly interpolated, it can cause errors as well.

Another issue was the sharp transition pattern between orbits that was observed in some of the features because of the significant time difference (approximately 20 days) in some orbits (Fig. 10). The data availability and acquisition time varied within a feature in Sentinel-1 orbits in this study. This could be a main obstacle to further improving the classification, and it is a common issue when a study area is large and data-acquisition times are far. If we were to classify orbit-wise, we would have insufficient field

reference for training and validation steps. Therefore, we had to mosaic the images before classifying. A solution can be to strip normalization between neighboring orbits, but it can cause artifacts. If the differences come from meteorological differences between acquisition times, it could be possible to generate a relatively smooth, eye-pleasing boundary by forcing some average backscatter to the same value on both sides of the boundary. This backscatter manipulation could then cause some targets away from the average backscatter to behave strangely. In addition, the balancing is different depending on the application area and the geographic area. If the area is covered by a forest, one may assume that the backscattering coefficient gamma-0 will remain constant, regardless of the incidence angle. This is not the case with many smooth (in terms of radar wavelength) agricultural targets.

3.3. Sensor comparisons

Comparing our achieved overall accuracy with World Bank (2016) that used multispectral sensor of Landsat-8, our results are justifiable and comparable by for example using dual-polarized SAR images, if available. During 2014-2015, almost all of the Sentinel-1 acquisitions over Pakistan were single polarized. Therefore, single vertical (VV) products were utilized for this study. Thus, we had 7 features and achieved 42% and 57% of overall accuracy for all and agricultural classes, respectively; unlike World Bank (2016) that had 35 spectral bands from Landsat-8 and yielded 59% and 68% of overall accuracy. The World Bank (2016) employed 5 bands of Landsat-8 that senses the visible, near-infrared, and short-wave infrared covering the spectrum range between 0.45–1.65 μm. However, the SAR has its assets: independent of sun light and cloud conditions.

3.4. Recommendations for future studies

Adding texture bands, using dual-polarized Sentinel-1 image (upon availability), and fusing with optical images, such as Sentinel-2 and Landsat-8 are recommended for future studies. Moreover, a time-series analysis is highly recommendable for selecting features more meaningfully. In addition, accounting for weather information (e.g., wind and rain) should be highly beneficial for several reasons. For example, rain increases the backscatter from all land cover types and lowers the image contrast, and wind causes turbulence on the surface of water and increases backscatter abnormally. Further, prior knowledge about the distribution patterns of a land cover and climate or microclimates within the large area should be substantially beneficial when mapping a land cover over large areas. Future studies could investigate methods that treat imbalanced training data. Moreover, it may be considered to interpolate the 1–2-pixel wide gaps properly or use SAR mosaicking or slice assembly functions in SNAP, which were not working properly during the data analysis of this research. Lastly, normalizing the sharp transition in pixel values in some orbits maybe considered. If available, the use of regional statistics and stratification along with Sentinel-1A SAR could improve the accuracy.

4. Conclusions

This study revealed the advantages and disadvantages of using Sentinel-1A SAR images in nation-wide land cover mapping. The study attempted to recover the weak point by using novel methods.

In this study, the Random Forest classifier outperformed the other tested machine-learning methods. We achieved a model with an overall accuracy of 43%, and the model was used to classify land cover. The overall accuracy of the final map was 42%. The accuracy for mapping agricultural classes was 57%. Our land use map was

promising for major classes, although differentiating small classes from each other was less promising. In line with the earlier Landsat study, the dominant classes were overestimated and rare classes were underestimated. This research note documented the land cover mapping using Sentinel-1 images for large-area, and the results were comparable with earlier studies in smaller areas and with less land cover classes. We made recommendations to achieve better accuracy in future mapping of large-area.

5. Acknowledgements

This study was a collaboration between Simosol Oy Ltd., the VTT Technical Research Centre of Finland, and the University of Helsinki. We would like to thank Simosol Oy, Ltd., for sponsoring this research and its software and hardware support. Mr. Hafiz Umair Masood Awan should be acknowledged for his constructive proof reading.

6. Declaration of interest

The authors declare no conflict of interest.

7. References

Abdikan, S., F. B. Sanli, M. Ustuner, and F. Calò. 2016. "Land Cover Mapping Using Sentinel-1 Sar Data." *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, no. June. doi:10.5194/isprsarchives-XLI-B7-757-2016.

Alexandrer, Jacob, and Ban, Yifang. 2015. "SENTINEL-1A SAR DATA FOR GLOBAL URBAN MAPPING : PRELIMINARY RESULTS." *Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International*, 1179–82. doi:10.1109/IGARSS.2015.7325982.

Balzter, Heiko, Beth Cole, Christian Thiel, and Christiane Schmullius. 2015. "Mapping CORINE Land Cover from Sentinel-1A SAR and SRTM Digital Elevation Model Data Using Random Forests." *Remote Sensing* 7 (11): 14876–98. doi:10.3390/rs71114876.

Dimov, D, J Kuhn, and C Conrad. 2016. "Assessment of Cropping System Diversity in the Fergana Valley Through Image Fusion of Landsat 8 and Sentinel-1." *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* III (July): 12–19. doi:10.5194/isprsannals-III-7-173-2016.

Gislason, Pall Oskar, Jon Atli Benediktsson, and Johannes R. Sveinsson. 2006. “Random Forests for Land Cover Classification.” *Pattern Recognition Letters* 27 (4): 294–300. doi:10.1016/j.patrec.2005.08.011.

Haas, Jan, and Yifang Ban. 2015. “Synergies of Sentinel-1a SAR and Sentinel-2a MSI Data for Urban Ecosystem Service Mapping.” In *35th EARSeL Symposium – European Remote Sensing: Progress, Challenges and Opportunities*. Stockholm, Sweden.

Imran, A, Q Zaman, G Rasul, and A Mahmood. 2014. “An Analytical Study of Variations in the Monsoon Patterns over Pakistan.” *Pakistan Journal of Meteorology* 10 (20): 25–37.

Kilsedar, Candan Eylül. 2016. “Sentinel Application Platform (SNAP) Basics.”

Makinde, Esther O, and Oluwaseun E Oyelade. 2018. “Land Cover Mapping Using Sentinel-1 SAR Satellite Imagery of Lagos State for 2017 †.” *Proceedings* 2: 1–5. doi:10.3390/proceedings2221399.

Nguyen, Duy Ba, Alexander Gruber, and Wolfgang Wagner. 2016. “Mapping Rice Extent and Cropping Scheme in the Mekong Delta Using Sentinel-1A Data.” *Remote Sensing Letters* 7 (12). Taylor & Francis: 1209–18. doi:10.1080/2150704X.2016.1225172.

Pal, Mahesh. 2003. “Random Forests for Land Cover Classification.” *Institute of Electrical and Electronics Engineers*, 3510–12.

Rodriguez-Galiano, V.F., Ghimire, B., Rogan J., Chica-Olmo, M., Rigol-Sanchez, J.P.. 2012a. “An Assessment of the Effectiveness of a Random Forest Classifier for Land-Cover Classification.” *ISPRS Journal of Photogrammetry and Remote Sensing* 67 (1). International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS): 93–104. doi:10.1016/j.isprsjprs.2011.11.002.

Rodriguez-Galiano, V.F., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P.M., Jeganathan, C.. 2012b. “Random Forest Classification of Mediterranean Land Cover Using Multi-Seasonal Imagery and Multi-Seasonal Texture.” *Remote Sensing of Environment* 121. Elsevier Inc.: 93–107. doi:10.1016/j.rse.2011.12.003.

Tapete, Deodato, and Francesca Cigna. 2016. “Urban Remote Sensing in Areas of Conflict: TerraSAR-X and Sentinel-1 Change Detection in the Middle East.” *CONFERENCE: Fourth International Conference on Remote Sensing and Geoinformation of the Environment* 9688: 968821. doi:10.1117/12.2241442.

Veci, Luis. 2015. “SENTINEL-1 Toolbox SAR Basics Tutorial.”

World Bank. 2016. “Biomass Resource Mapping in Pakistan: Final Report on Biomass Atlas.” Energy Sector Management Assistance Program. Washington, D.C. http://documents.worldbank.org/curated/en/104071469432331115/Biomass-resource-mapping-in-Pakistan-final-report-on-biomass-atlas.

437 ========== List of Tables. ==========

438 Table 1. Sentinel-1a images information, used for land use classification.

epoch (feature)	Time limit from	Time limit to	range (number of days)	number of images
1	2015-03-28*	2015-04-26*	29	34
2	2015-05-15*	2015-06-06*	22	35
3	2015-06-05*	2015-07-07*	32	33
4	2015-09-09*	2015-10-08*	29	34
5	2014-10-13	2014-11-14	32	34
6	2014-11-09	2014-12-01	22	34
7	2015-03-04	2015-03-26	22	35

439 *Refers to time limits that mismatches with the time of field data collection, the rest

440 matches.

441

442 Table 2. The time limits for Landsat 8 features used in land use classification (source:

443 World Bank (2016)).

Epoch (feature)	Time limit from	Time limit to
1	2014-04-01	2014-05-06
2	2014-05-14	2014-06-02
3	2014-06-04	2014-06-29
4	2014-09-08	2014-09-29
5	2014-10-03	2014-10-22
6	2014-11-11	2014-11-25
7	2015-03-08	2015-04-22

444

445 Table 3. List of land use classes in ground reference data (source: (World Bank 2016)).

Field classes (Kharif-Rabi)	Other classes
cotton-maize	bare area
cotton-other crop	low biomass
cotton-rice	medium biomass
cotton-sugarcane	high biomass
cotton-wheat	road
maize-maize	urban
maize-rice	water
maize-sugarcane	
maize-wheat	
other crop-other crop	
rice-rice	
rice-sugarcane	
rice-wheat	
sugarcane-sugarcane	
sugarcane-wheat	
wheat-wheat	

446 Table 4. Number of field observation data used for training and validation per land

447 cover type.

	Number of field data in	
Land cover type	**training**	**validation**

cotton-maize	18	7
cotton-other crop	11	4
cotton-rice	7	6
cotton-sugarcane	17	8
cotton-wheat	944	243
maize-maize	62	16
maize-rice	14	5
maize-sugarcane	17	6
maize-wheat	111	35
Other crop-other crop	70	20
rice-rice	56	16
rice-sugarcane	15	5
rice-wheat	735	200
sugarcane-sugarcane	97	29
sugarcane-wheat	199	51
wheat-wheat	112	25
bare area	192	49
high biomass	530	142
low biomass	460	109
medium biomass	629	149
road	52	20
urban	136	41
water	141	31
Total	4625	1217

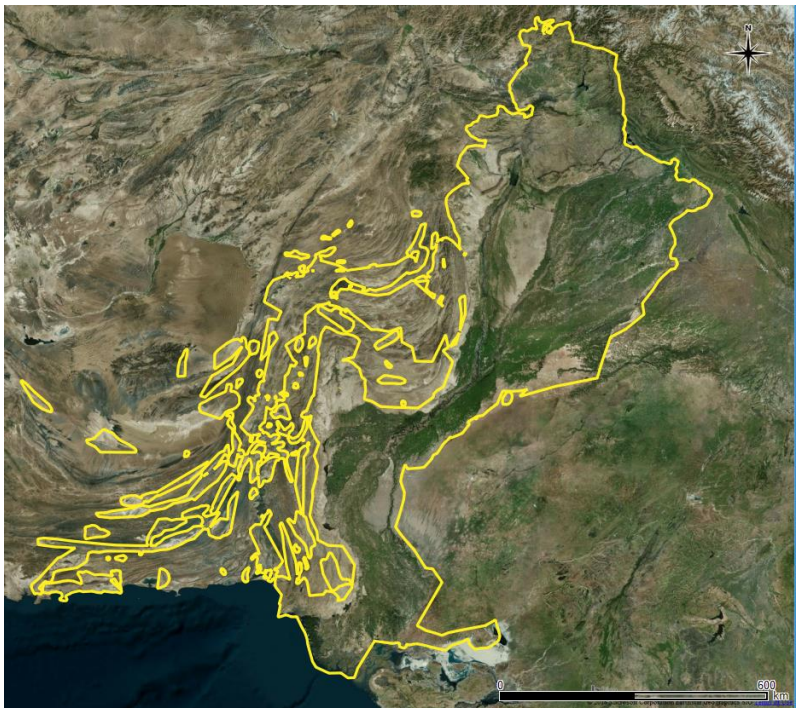
448

449 Table 5. The modelling accuracy of all seven machine learning methods.

Method	Overall Accuracy (%)
AdaBoost	31
GaussianNB	19
KNeighbors	37
Linear Discriminant Analysis	25
Quadratic Discriminant Analysis	22
Random Forest	43
Support Vector Machine (SVC module)	14

450

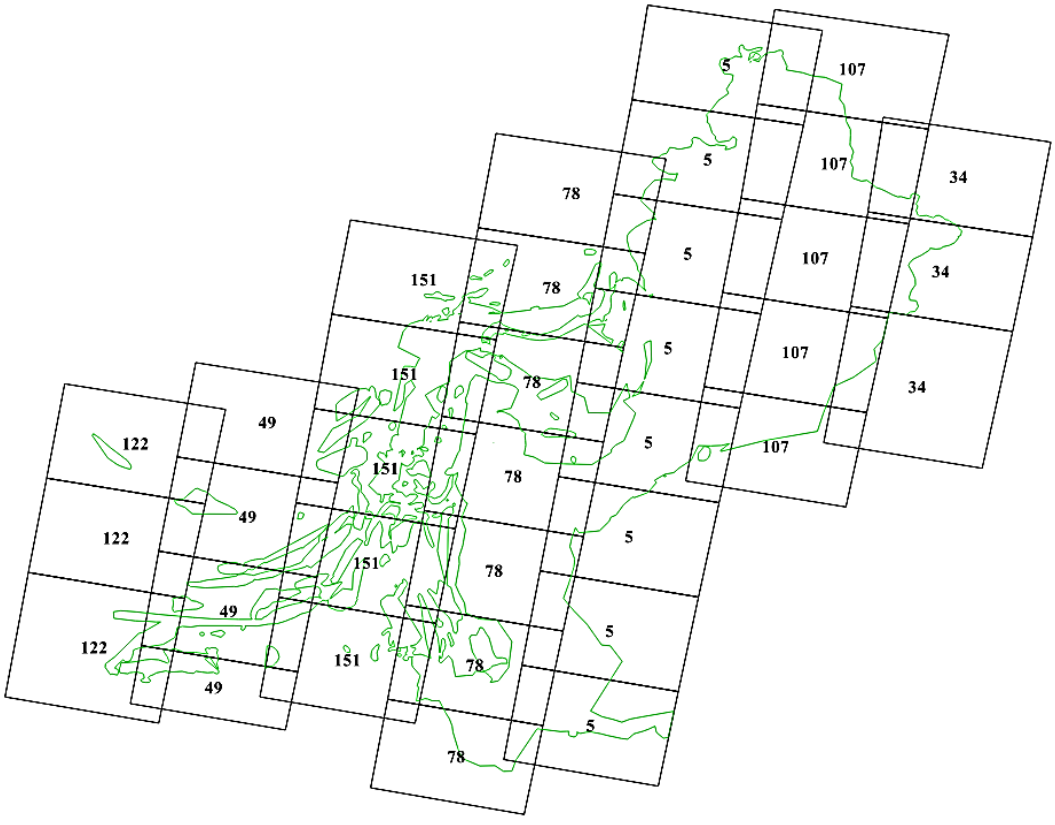
451 ========== List of Figures ==========



452

453 Fig. 1 The study area (yellow) visualized with background Bing Aerial photo.

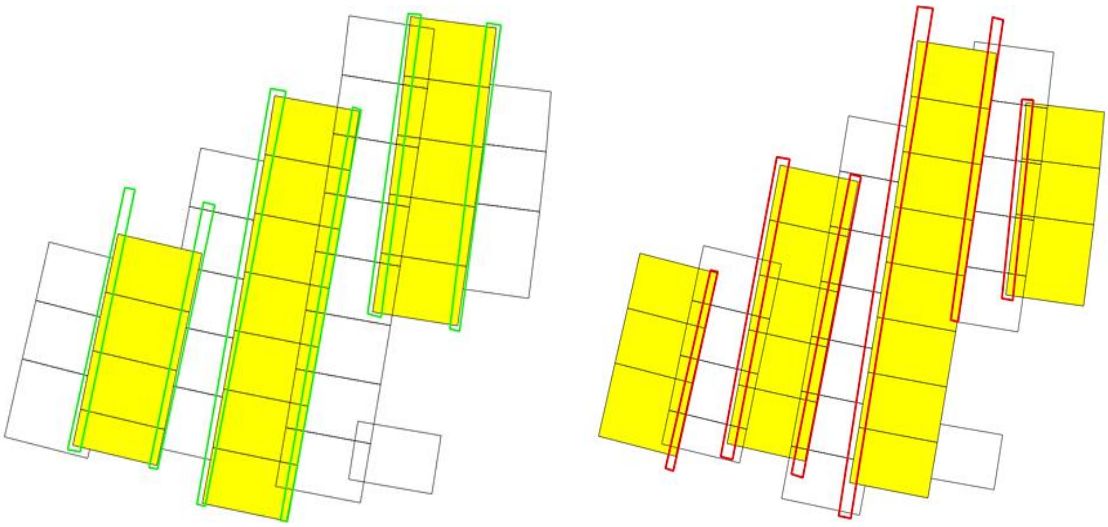
454



455

456 Fig. 2 Sentinel-1A images used in one feature (epoch) with their relative orbit number.

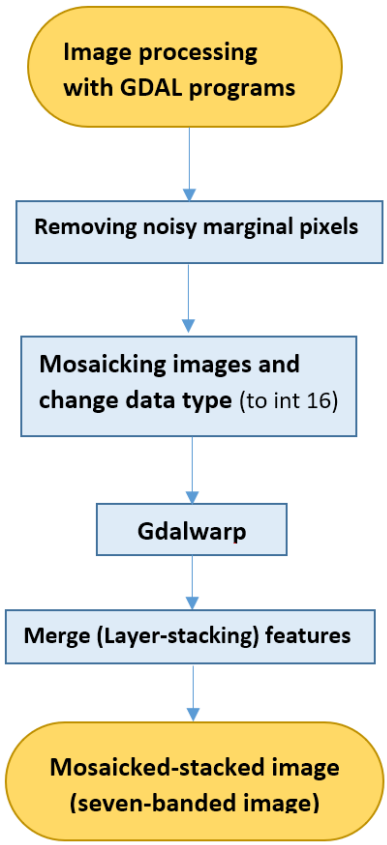
457



458 Fig. 3 Illustration of cut shapefile (green in left and red in right image) and yellow

459 highlighted orbited image to cut.

460



461

462 Fig. 4 Image processing in GDAL, executed after processing of SAR data in SNAP.

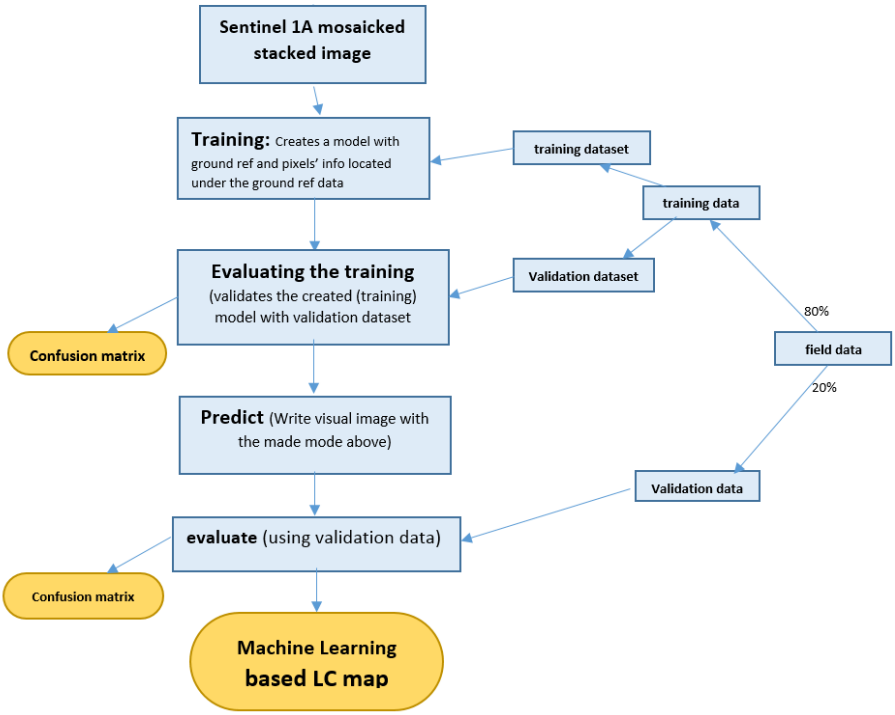


Fig. 5 Simplified flowchart of Machine Learning based method.

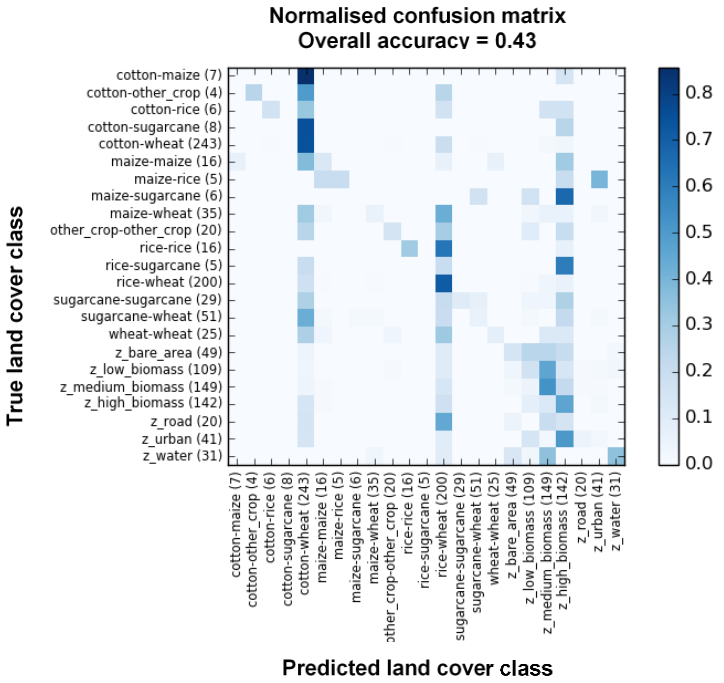
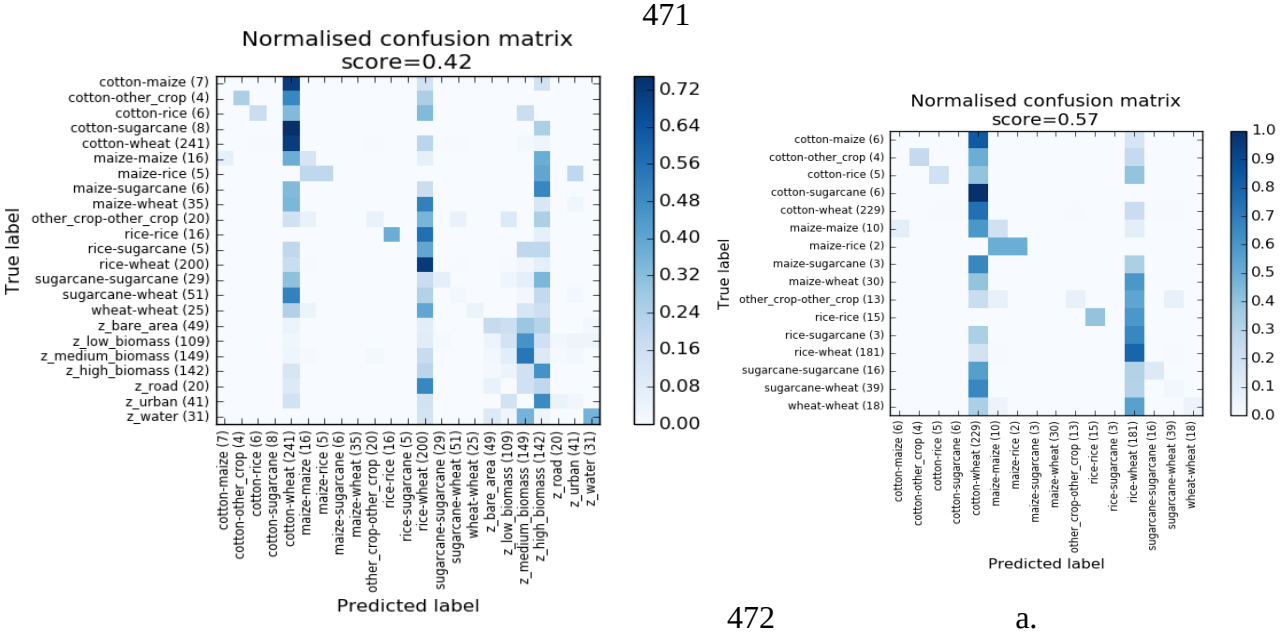


Fig. 6. The normalized confusion matrix for Random Forest modelling step. The numbers in parenthesis indicate the number of classes as in the true or predicted land

470 cover classes.

471



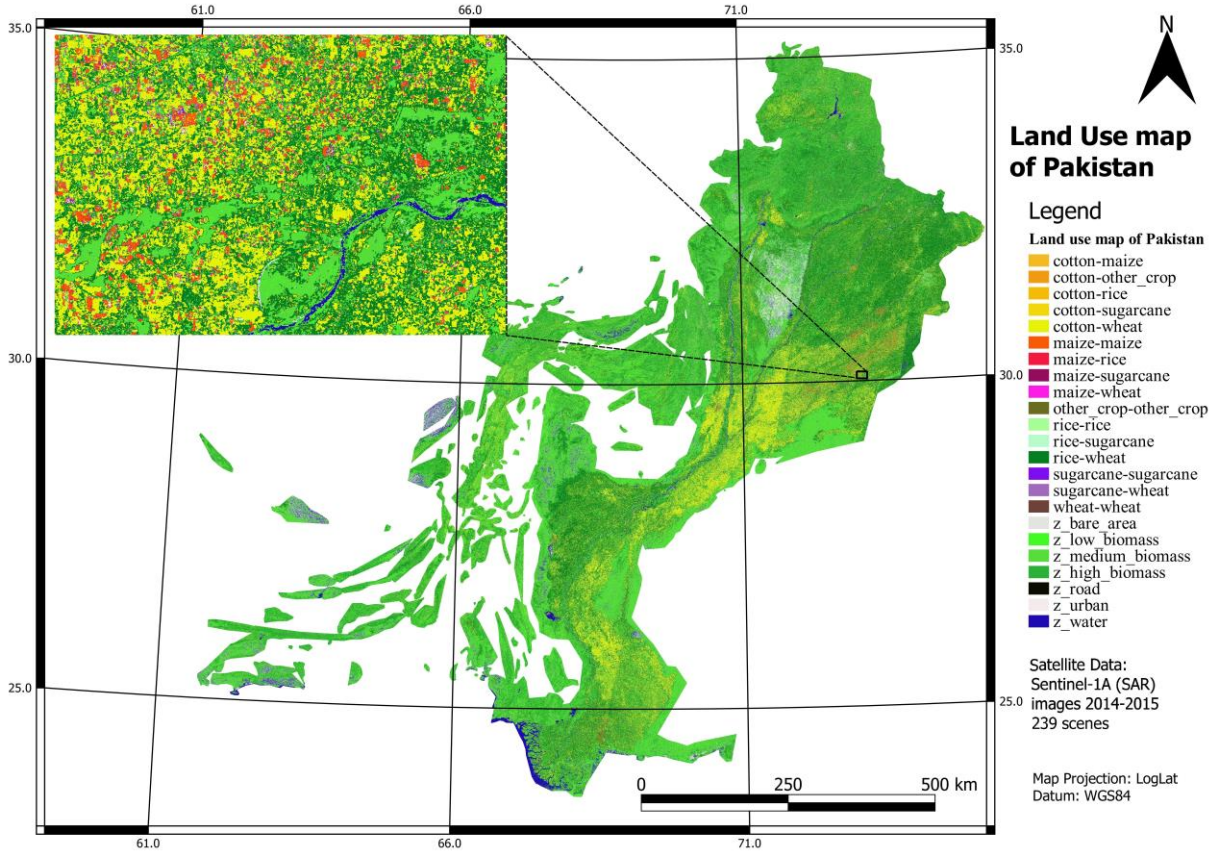
472 a.

473 b.

474 Fig. 7. Land cover classification confusion matrix for Random Forest validation step for

475 all classes (a) and for agricultural classes (b). The numbers in parenthesis indicate the

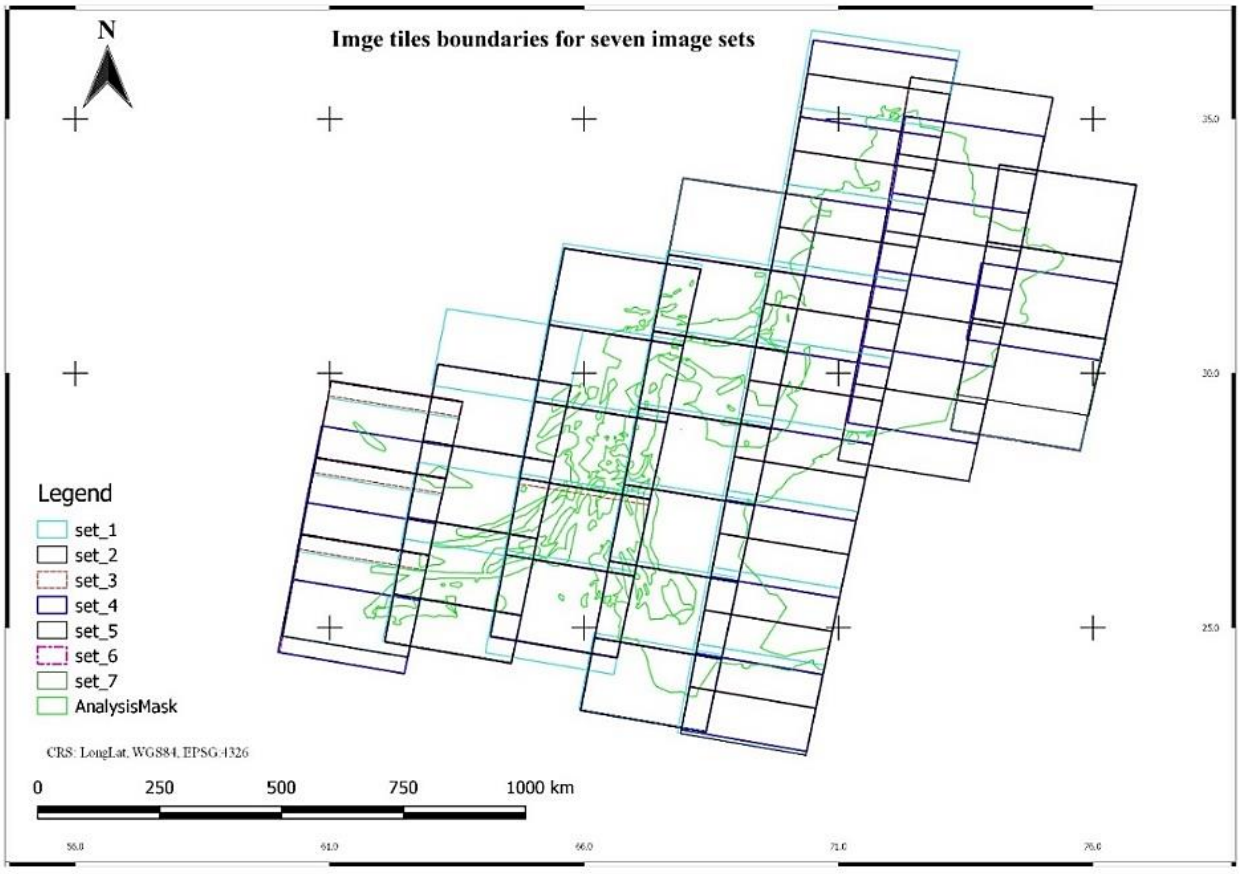
476 number of classes as in the true or predicted land cover classes.



477
478 Fig. 8. Land cover map of Pakistan using Sentinel-1A SAR images (date: 2014-2015,

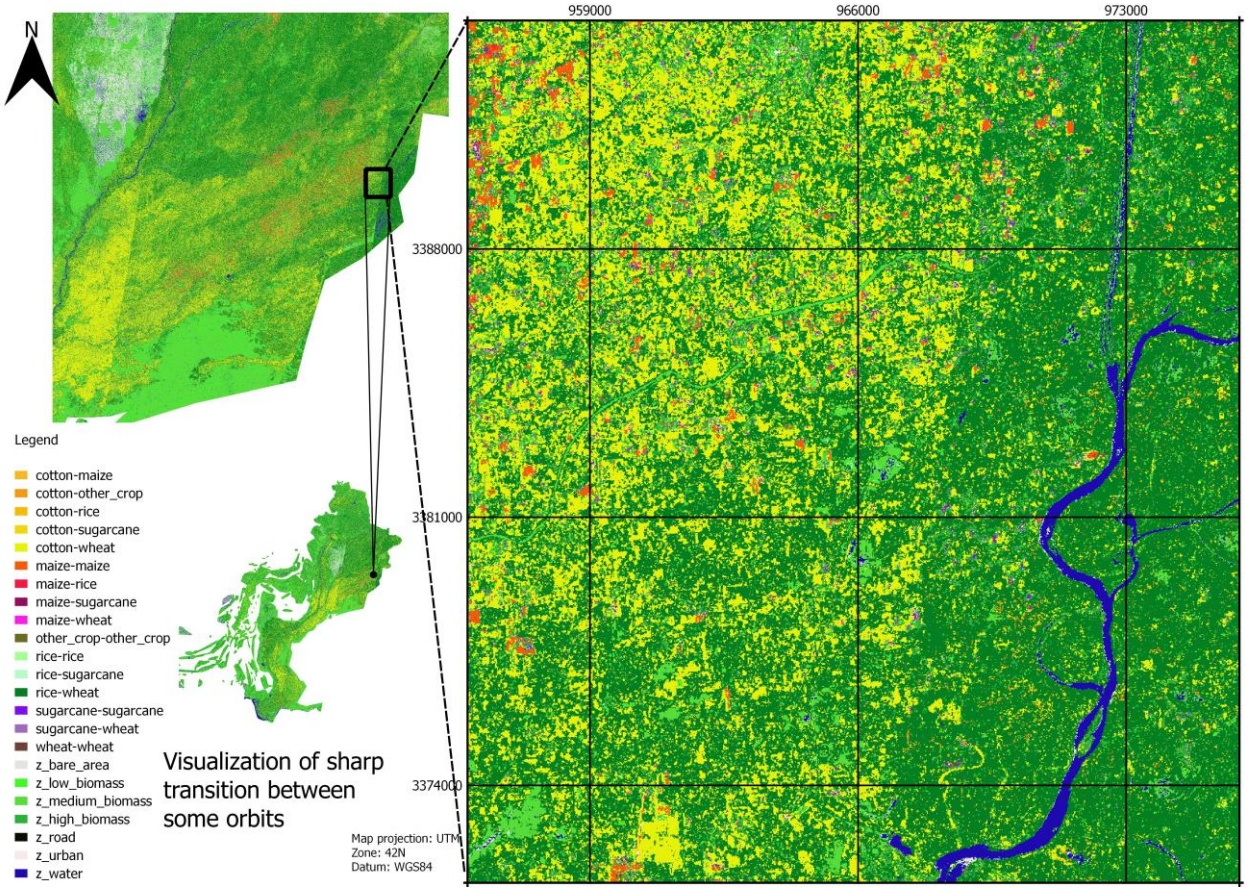
479 total number of images used: 239).

480



481 Fig. 9. Map of images boundary in each feature (set) showing the inconsistency in the

482 image's boundaries especially in feature 1 and 3.



483

484 Fig. 10. Sharp transition in some far-date orbits (20 days) caused misclassification of

485 Sentinel-1A images (date: 2014-2015).